Watching Closely

Bálint, Katalin; Klausch, Thomas; Pólya, Tibor

published in
Journal of Media Psychology
2016

DOI (link to publisher)
10.1027/1864-1105/a000189

document version
Peer reviewed version

Original Article

Watching Closely

Shot Scale Influences Theory of Mind Response in Visual Narratives

Katalin Bálint,[1] Thomas Klausch,[2] and Tibor Pólya[3]

[1]Department of Media, Knowledge and Communication, University of Augsburg, Germany
[2]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands
[3]Károli University of the Reformed Church, Institute of Psychology, Budapest, Hungary

Abstract: Theory of mind, defined as attributing intentions, feelings, and thoughts to others, is a key capacity in building mental models of narratives. Shot scale, referring to the relative size of the figure on the screen, effectively regulates theory of mind relevant visual cues carried by faces. However, research into the effect of shot scale on theory of mind responses in a narrative is almost nonexistent. The aim of the present study was to investigate the extent to which shot scale influences theory of mind responding in film viewers. Four short animated movies were annotated for average shot scale and presented in a within-subject design. Employing mixed-method data collection, participants were asked to retell the story of the films and fill in questionnaires on narrative experience. Skin conductance was also measured during exposure. Story descriptions were content analyzed for theory of mind responses. In a Poisson regression model, average shot scale predicted theory of mind response indicating that increasing spatial proximity to the character triggered higher occurrence of mental state references in participants' story descriptions. The findings elucidate how formal properties of character presentation affect an audience's mental models of a story.

Keywords: shot scale, theory of mind, spatial distance, character engagement, visual narratives

Theory of mind (TOM), defined as the awareness of mental states in others (Gallagher & Frith, 2003), has a crucial role at every level of human sociality (Heyes & Frith, 2014). Understanding the mental states in others is a key component of interpersonal interactions (Epley & Waytz, 2010). People are constantly faced with situations in which they have to recognize what the other person feels, wants, or knows in order to effectively react to or predict the other person's behavior. TOM has a high social value, and it is thus important to elucidate what media-specific features can influence this response. A growing body of research showed that both literary (Kidd & Castano, 2013) and cinematic (Black & Barnes, 2015) fictional narratives have a huge potential to elicit and improve TOM responding. However, these studies focused mainly on the broad comparison of fiction and nonfiction narratives, and neglected the specific role of non-content-related audiovisual form features. To address this research gap, the present study investigates the effect of one specific form feature, shot scale, on TOM responding in viewers. The findings of this study may provide media producers and scholars in entertainment-education with empirically tested information to develop effective audiovisual messages for eliciting TOM responding. The following section clarifies the role of TOM in empathy responding and character identification, and then it introduces previous research into the effect of shot scale and spatial distance.

TOM has a strong correspondence to other social cognition processes, such as empathy (Völlm et al., 2006). However, there has been no final agreement in the literature about how these partially overlapping constructs should be separated (Decety & Ickes, 2009; Decety & Jackson, 2004). In recent social cognitive neuroscience, two systems of empathy have been differentiated (Dziobek et al., 2008; Lieberman, 2007; Raz & Hendler, 2014; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). TOM is closely associated with cognitive empathy, that is, representing the psychological states of others, which is different from affective empathy, that is, embodied simulation or experiencing the mental state of others (Lieberman, 2007). In previous studies TOM has been considered either as a trait-like capability that can be impaired in various psychological disorders (e.g., autism spectrum disorder; Baron-Cohen, 2001; borderline personality disorder; Fonagy & Bateman, 2007; schizophrenia; Corcoran, Mercer, & Frith, 1995) or as a state-like response that may or may not be activated in the context of a certain relationship (Meins, Fernyhough, & Harris-Waller, 2014). It seems that a person's tendency to be aware of the mental states of others does not only depend on the individual capability of the observer, but also

on the characteristics of the observed person and the nature of the relationship between them (Haslam, 2006). For example, members of stigmatized groups are usually ascribed only a few and simplistic mental states by others, and this so-called dehumanization has profound consequences on their lives (Epley & Waytz, 2010).

TOM is also crucial in audience members' engagement with fictional characters. Cohen's well-known conceptualization of identification comprises the dimensions of loss of self-awareness and sharing the feelings, motivation, and perspective of the character (Cohen, 2001). TOM function is required for this latter perspective-taking dimension that can be measured by the degree to which "an audience member feels he or she understands the character and the motivations of his or her behaviour" (Cohen, 2001, p. 256). Recently, it has been recognized that perspective taking is of special significance in viewers' social acceptance of fictional characters from stigmatized groups (Chung & Slater, 2013). It seems that TOM is one of the underlying psychological mechanisms of identification that contributes to the prosocial impact of narratives.

TOM is also an important building block of the experience with mediated narratives (Levin, Hymel, & Baker, 2013; Raz & Hendler, 2014). The event-indexing model of story comprehension explains that when exposed to narratives, audience members construct a primary mental model of the story based on recognition of events and *intentional actions* of characters (Zwaan, Langston, & Graesser, 1995). This indicates that a mental representation of a story requires the awareness of the characters' intentions (Zwaan et al., 1995). In line with this, research showed that brain areas of TOM are activated when people are exposed to narratives (Mar, 2011). In summary, TOM can be regarded as one of the basic ingredients of character engagement (through empathy or identification) and narrative understanding rendering the narrative experience more enjoyable and relatable.

Based on findings in neuroscientific research, it is reasonable to assume that shot scale may influence TOM responding. Shot scale, defined as the apparent spatial distance of characters from the camera, is one of the most effective visual devices with which to regulate the relative size of characters' faces, the relative proportion of the human figure to the background (Bowen & Thompson, 2013; Salt, 1992), as well as to arrange film content according to its saliency (Carroll & Seeley, 2013). Relatedly, TOM response is strongly associated with visual cues carried by human faces (Calder et al., 2002; Itier & Batty, 2009; Itier, Villate, & Ryan, 2007; Mosconi, Mack, McCarthy, & Pelphrey, 2005). Processing facial expressions and gaze direction are salient triggers of TOM (Frischen, Bayliss, & Tipper, 2007), likely because the eye region conveys salient information about emotion (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Calder et al., 2002), face recognition (Hood, Macrae, Cole-Davies, & Dias, 2003), and the target of attention (Frischen et al., 2007). Thus, it is reasonable to assume that shot scale, in being able to regulate audiences' visual access to fictional characters' faces, may influence TOM responding (see Levin et al., 2013).

The family of basic shot scales includes long shot, medium shot, and close-up shot. In long shots the human figure is small and surrounded by a large area of the film space. The medium shot depicts the human figure from the waist up corresponding to a distance of 3–5 feet. The close-up shot shows the figure from the upper shoulder, and provides a detailed view of the human face (Bowen & Thompson, 2013). Previous research has shown that closer shots increase arousal (Canini, Benini, & Leonardi, 2011), evoke empathic care (Cao, 2013), memory (Mutz, 2007), and intensify liking/ disliking of a character (Mutz, 2007). Yet, none of these previous studies reflected on TOM. However, it is noteworthy that in Cao's (2013) study, empathy was measured by items focusing on perspective-taking (e.g., "I imagined what it was like to be in his situation.") which has a strong link to TOM responding. These results indicate that shot scale carries socially important information and shapes engagement with media characters. In line with this assumption, the power of actors' faces to engage viewers has been widely discussed in film theory (Balazs & Carter, 2013; Carroll, 1993; Persson, 2003; Plantinga, 1999). It has been argued that close-up shots elicit empathic emotions (Plantinga, 1999) and attribution of mental states to characters (Carroll, 1993; Persson, 2003; Smith, 1995), although empirical evidence of this is still limited. Despite the obvious theoretical link between shot scale and TOM, there have been hardly any empirical studies focusing on their relationship. Addressing this research gap, the present study examines the effect of average shot scale of narratives on viewers' TOM responding.

Shot scale, owing to shared mechanisms of depth perception (Howard, 2012), is closely related to other spatial distance properties (e.g., viewing distance, image size, and screen size) of media messages (see Detenber & Lang, 2010). Spatial distance – the perceived physical distance between the observer and the observed – has been shown to have a significant role in our real-life experiences (see, e.g., Hall & Hall, 1969), as well as in media experience (Lombard, 1995). Processing spatial distance (similarly to shot scale) is a subject of the size–distance law: The perceived image size shrinks as the object moves farther away (Howard, 2012). According to Detenber and Lang (2010), larger images increase automatic allocation of cognitive capacity, require more visual processing, and consequently generate larger mental images. In line with this model it has been shown that decreased apparent distance increased physiological arousal (Codispoti & De Cesarei, 2007; De Cesarei & Codispoti, 2010), self-reported arousal

(Detenber & Reeves, 1996), attention (Reeves, Lang, Kim, & Tatar, 1999), memory (Bellman, Schweda, & Varan, 2009), emotions (De Cesarei and Codispoti, 2006, 2008), immersion (Baranowski & Hecht, 2014), and presence (Hou, Nam, Peng, & Lee, 2012; Lombard, Ditton, Grabe, & Reich, 1997; Lombard, Reich, Grabe, Bracken, & Ditton, 2000). Important to the purpose of the present study, is the finding that spatial distance also has an impact on character engagement. Shorter apparent distance increases liking (Bellman et al., 2009; Hou et al., 2012; Lombard, 1995), and the intensity of the perceived emotions of a character (Lombard et al., 1997). However, TOM has not been reflected in these studies. Similarly, looming (i.e., camera zooming) increases arousal (Mühlberger, Neumann, Wieser, & Pauli, 2008), cognitive load (Lang, Bradley, Park, Shin, & Chung, 2006), and attention (Franconeri & Simons, 2003). In summary, in mediated environments closer figures trigger more intense reactions (Lombard, 1995; Reeves & Nass, 1996).

According to the limited capacity for mediated processing (LC4MP) model (Lang, 2000), varying the type and number of structural features in a message influences motivational activation and audiences' cognitive and emotional responses (see Detenber and Lang, 2010). The model claims that even though viewers are able to control their attention to stimuli based on their own intentions, formal features of messages can still elicit automatic involuntary attentional processes. These attentional processes then activate motivational systems (i.e., approaching or avoiding tendencies), and influence cognitive (e.g., storage and retrieval) and emotional reactions. Following this line of argument, we assume that shot scale may increase the motivational relevance of the image by bringing the character closer to the viewer, increasing the size of the face, thus making the details of facial expression more visible and readily available. The heightened motivational relevance increases the allocation of additional resources, which may raise the probability that the characters' mental states become part of the mental model of the story created by the viewer. As a consequence, we assume that the viewers' model of the narrative will be more mentalized, that is, the narrative plot will be remembered with more references to the mental state of the characters. To put it differently, we hypothesize that decreased spatial distance from the character will be associated with higher occurrence of TOM responding.

Method

Overview

Four animated films were annotated for shot scale in order to calculate an average shot scale measure. The one-factorial within-subjects design called for each participant to view three of four animated films (owing to time constraints) while skin conductance was recorded. Immediately after viewing each of the films, participants rated their identification with the characters, emotional valence, narrative understanding, perceived reality, and enjoyment. After watching all films, participants were asked to retell the story of the movies seen. The occurrence of TOM references was counted in these film-based narratives. A Poisson regression analysis was performed to test the influence of average shot scale on the occurrence of TOM. The duration of the films, number of coding units, skin conductance, identification, emotional valence, narrative understanding, perceived realism, and enjoyment were included as control variables in the analysis. Having each participant in each average shot scale condition made it possible to separate treatment effects and between-subject variance. Thus heterogeneity in TOM ability, the ability to verbalize mental states of characters, and gender effects could be controlled (Reeves & Geiger, 1994).

Participants

The study employed convenience sampling. We recruited 60 participants using university e-mail lists, social media platforms, and psychology-related public homepages. We excluded five participants because they had already seen one of the films projected, and three participants because of technical problems during data collections. The final research sample consisted of 52 participants (42 female) aged between 18 and 60 years (M = 25.6, SD = 6.7). University students made up 65% of the sample, and35% were in employment. We rewarded participants with a book. All participants gave informed consent for their participation in the experiment.

Stimulus Materials

We selected four films that represent different levels of average shot scale from a large collection of awarded short animation films. Besides average shot scale, the systematic selection was guided by knowledge of other message features that could influence TOM, and it aimed at minimizing potential confounds. The following films were used in the present study: *Father and Daughter* (de Wit, 2001); *Invention of Love* (Shuskhov, 2010); *Lavatory – Lovestory* (Boyarskiy, Selyanov, & Bronzit, 2007); *Lettingo* (Whitaker, 2008). In order to reduce the potential confounding variables in the study, we attempted to select films that were similar to each other at many levels. All four films featured a self-contained fictional narrative with a linear narrative structure presenting a separation – reunion story of a female

and a male character. The female character had a major role, whereas the male character had a minor role in the films. Importantly, all four films presented anthropomorphic protagonists in realistic style. The films were dialogue-free (i.e., no voice-over narration, textual inserts, or lyrics), and the images were combined with a continuous extra-diegetic instrumental music. The length of the films was between 4.67 and 8.9 min. Several aspects are worth noting with regard to the stylistic features of the films. First, all four movies were two-dimensional animations, one of them was digital (*Invention of Love*) and three of them were hand-drawn animations. All four films used a simplistic style characterized by a very limited color palette (black and white, monochromatic, or two to three colors) and simple, clean lines. Three movies of the four (except the film *Lettingo*) were awarded multiple international awards for their artistic merits.

For each movie we calculated an average shot scale value that expresses the average distance of the character from the camera throughout the whole film. The procedure was as follows. In the first step, movies were segmented into shots (i.e., segments between two cuts), and the shots presenting a character were selected and their duration was measured (*Lavatory – Lovestory* 5.25 min; *Father and Daughter* 5.02 min; *Invention of Love* 4.58 min, *Lettingo* 2.2 min). Next, the selected shots were annotated manually by shot scale (extreme long shot, very long shot, long shot, medium long shot, medium shot, medium close-up, close-up, big close-up, extreme close-up). The duration of shots in the same shot scale was aggregated. In the next step, the percentage of aggregated shot scale durations in the total duration of shots presenting a character was calculated. These percentages of shot scale duration were multiplied by the weight assigned to shot scales. Higher weights expressed higher distance (extreme long shot = 9, very long shot = 8, long shot = 7, medium long shot = 6, medium shot = 5, medium close-up = 4, close-up = 3, big close-up = 2, extreme close-up = 1). In the last step, the final value of average shot scale for each movie was calculated by summing the weighted shot scale proportions (*Father and Daughter* = 8.22; *Invention of Love* = 6.10; *Lettingo* = 5.47; *Lavatory – Lovestory* = 5.02). Higher values indicated a higher average distance of the character from the camera, that is, larger shot scales in general.

Procedure

Participants completed the experimental protocol individually. After a welcome, the participant was led into the laboratory and seated in front of a computer. Viewing distance averaged approximately 50 cm. Participants were initially informed that the purpose of the study was to determine how people process fictional narratives and that it would take 65 min, in which the participant would watch three short animation films. It was explained that skin conductance would be recorded using small sensors while watching. The participant signed a consent form and two finger straps were placed against the inside part of the index and ring finger of the nondominant hand. Participants were asked to sit quietly and to fully watch each video clip. During the experiment, skin conductance data were collected just prior to the onset of each film for 1 min (for the baseline data) as well as for the duration of the film. Participants were shown the first film clip and were asked to fill in questionnaires about the film experience. This process was repeated for three of the film clips, which were presented in a counterbalanced order across participants. To keep response burden manageable, each participant viewed only three of the four movies. All participants viewed a film with a high (*Father and Daughter*) and medium level (*Invention of Love*) of average shot scale, and subsequently a film with a low average shot scale (either *Lettingo* or *Lavatory – Lovestory*). The latter assignment was made at random. After having seen all three films, participants were asked to: "Tell about what happened in the film scenes you just saw." Participants were asked to recall the three film scenes in a randomly assigned order. Their verbal responses were recorded. Finally, an open question addressed the participants' awareness of the hypothesis, and the result of this showed that they were not aware of the hypothesis. The session was closed with debriefing and rewarding.

Software and Equipment

Films and questionnaires were presented by MediaLab v2010 software (Empirisoft Corporation), which enabled us to systematically vary the successive phases of the study. We used an LCD monitor (HP L1908, 19-inch, 446 × 59 × 301 mm, 1,440 × 900 pixels) and two 0.6-Watt surround-sound speakers for film projection. Verbal data were recorded on a digital voice recorder (Olympus VN-8500PC). Data management and analysis were performed using Matlab and R.

Measures

Theory of Mind

Each participant's set of three film-based narratives was coded for the occurrence of TOM response. TOM response was assessed by counting the frequency of mental state references in the coded transcripts of the film-based narratives, that is, answers to the question to describe the story of the films. The counting procedure was informed by standardized assessments of TOM using story-based stimuli

and qualitative data collection (Barnes, Lombardo, Wheelwright, & Baron-Cohen, 2009; Dodell-Feder, Lincoln, Coulson, & Hooker, 2013; Dziobek et al., 2006; Golan, Baron-Cohen, Hill, & Golan, 2006; Heavey, Phillips, Baron-Cohen, & Rutter, 2000). The procedure was as follows: Participants' film-based narratives were transcribed verbatim, then segmented into subject–verb–predicate coding units and coded by two independent coders. A coding unit was defined as a word or string of words identified by a pause or grammatical completeness. Units were coded for the presence or absence of an explicit reference to a mental state of the main protagonist (i.e., the female protagonists). The coding was informed by Meins and Fernyhough's (2010) coding manual, in which mental reference is defined as any reference to an individual's mental life, relating to desire, wish, emotion, will, mind, imagination, interest, intellect, or metacognition. Coding was done by two trained independent coders blind to the hypothesis of the study. Interrater reliability was assessed in each film separately (Freelon, 2010) and found to be acceptable (Cohen's κ = .71 to .76). Disagreements between coders were resolved by a third coder serving as a tiebreaker (Lombard, Snyder-Duch, & Bracken, 2002). Participants were given a score on each film-based narrative for the count of the coding units containing an explicit mental reference. Higher counts are indicative of increased activity of TOM, and a more mentalized mental representation of the film. (In the present sample: M = 0.12–0.30; SD = 0.09–0.14.)

Assessing TOM responses by nondirective open-ended questions enabled us to obtain the spontaneously occurring TOM responses. By contrast, Likert-type tests (e.g., Kozak, Marsh, & Wegner, 2006) and other performance tests (e.g., Dziobek et al., 2006; Golan et al., 2006) often rely on forced choices and prompt mental reasoning directly (e.g., "What does the character feel?"); consequently, these tests are less sensitive to the absence of mental state references, and are less valid representations of individual differences in adults' spontaneous TOM abilities (Meins et al., 2014).

Identification

In order to control for the different levels of character engagement we included five items measuring respondents' felt emotional proximity with characters. The items were taken from Tal-Or and Cohen's (2010) identification scale (e.g., "While viewing, I felt like the main character felt."). Respondents indicated the extent to which they agreed with each statement on a 7-point scale, ranging from 1 (= *strongly disagree*) to 7 (= *strongly agree*). The ratings for the individual items were averaged, with higher scores indicating higher emotional proximity to the fictional character. Reliabilities were good for all clips (Cronbach's α = .82–.88; M = 4.40–5.37; SD = 1.08–1.35).

Affective Valence

In order to control for the affective valence of the experience, we included one item taken from the Self-Assessment Manikin scale (Bradley & Lang, 1994). Respondents rated the valence of their feelings caused by the movie on a 7-point scale, ranging from −3 (= *very negative*) to +3 (= *very positive*). For the later analysis, ranging was recoded to a more conventional 1–7 scale. Higher scores indicate more positive affective valence (M = 4.08–4.81; SD = 1.63–1.96).

Enjoyment

In order to control for the possible quality differences among movies, we included an enjoyment scale with three items (Tal-Or & Cohen 2010), such as: "I enjoyed the film segments I watched in the experiment very much," and "This is a movie that I can enjoy." Respondents indicated the extent to which they agreed with each statement on a 7-point scale, ranging from 1 (= *strongly disagree*) to 7 (= *strongly agree*). The ratings for the individual items were averaged, with higher scores indicating higher enjoyment of the film Reliabilities were good for all clips (Cronbach's α = .90–.92; M = 4.40–5.54; SD = 1.23–1.92).

Perceived Realism

Perceived realism was assessed by three items based on Tal-Or and Cohen (2010). The items were adjusted to the theme of the films in the present study. Respondents indicated the extent to which they agreed with each statement (e.g., "The events in the scenes resemble ones in [the] real world.") on a 7-point scale, ranging from 1 (= *strongly disagree*) to 7 (= *strongly agree*). The ratings for the individual items were averaged, with higher scores indicating higher perceived realism of the film. Reliabilities were good for all clips (Cronbach's α = .76–.89; M = 4.95–5.44; SD = 1.13–1.43).

Narrative Understanding

Three items were taken from the narrative engagement scale (Busselle & Bilandzic, 2009; e.g., "My understanding of the characters is unclear."). Respondents indicated the extent to which they agreed with each statement on a 7-point scale, ranging from 1 (= *strongly disagree*) to 7 (= *strongly agree*). The ratings for the individual items were averaged, with higher scores indicating better understanding of the narrative. Reliabilities were good for all clips (Cronbach's α = .83–.95; M = 4.06–6.32; SD = 0.79–1.37).

Arousal

Autonomic arousal was measured by the changes in the tonic level of electrical conductivity of the skin (skin conductance level). Physiological data were collected with a transportable measurement system BioGraph Infiniti (SA7900, Version 5.1.0, Thought Technology Ltd.) and

was processed by Ledalab Software, using adaptive smoothing and continuous decomposition analysis (Benedek & Kaernbach, 2010) that decomposes skin conductance (SC) data into continuous tonic and phasic activity. This method was based on standard deconvolution and was comparatively fast and quite robust. Sampling frequency was 256 Hz, which was down-sampled to 16 Hz in Ledalab. Change scores were calculated for each movie by subtracting the mean tonic activity for the baseline period prior to the movie from the mean tonic activity for the film viewing period. (In the present sample: $M = -0.13$–0.01; $SD = 0.57$–0.84.)

Analytic Strategy

To assess the effect of average shot scale on the count amount of TOM occurrences, we used a random intercept Poisson regression model, in which three measurement occasions of TOM were nested in each subject. Including the random intercept allowed us to control for potential heterogeneity in TOM ability across subjects and thus correlated TOM measurements within subjects. In addition, we took into account that the number of coding units resulting from the content analysis varied over subjects and movies depending on the length and level of detail of subjects' responses. Since the number of coding units determined the theoretical *maximum* of occurrences of TOM, the log of coding units was introduced as an *offset* into the model (Agresti, 2003). The final model estimates the effect of average shot scale on the *rate of occurrence* of TOM.

Results

We analyzed the data in two steps (Table 1). First, we estimated the effect of average shot scale on TOM controlling for the duration of the movies (Model 1). Second, we introduced the z-standardized subject-level covariate variables (identification, affective valence, narrative understanding, perceived realism, arousal, and enjoyment) into Model 1 leading to Model 2. This step controlled for subject-level variables as potential alternative explanations for TOM. For example, based on the literature of character identification, it could be a reasonable alternative explanation that participants who identified more with the characters generated more TOM in their storytelling independently of shot scale.

The results showed that, while controlling for duration, average shot scale had a negative effect on TOM in respondents' stories, suggesting that the expected rate of TOM in respondents decreased by factor $\exp(-.2438) = .7836$ with each additional unit of average shot scale ($p < .001$, Model 1). In addition, this model revealed that the length of a movie (duration) had a positive effect on TOM ($p < .001$).

Table 1. Random effect Poisson regression models of the effect of average shot scale on theory of mind response (controlled for duration and z-standardized covariates)

	Model 1	Model 2
Intercept	−0.9037 (.288)**	−0.6088 (.372)
Average shot scale	−0.2438 (.033)***	−0.2529 (.037)***
Duration	0.0998 (.026)***	0.0675 (.035)$^{+}$
Identification	–	−0.0335 (.058)
Affective valence	–	−0.0598 (.055)
Enjoyment	–	−0.0193 (.069)
Perceived realism	–	−0.0031 (.057)
Narrative understanding	–	0.1027 (.067)
Arousal (skin conductance)	–	0.0032 (.041)
Random effect variance	.0923	.0938
AIC	726.0	733.3

Note. AIC = Akaike information criterion. $^{+}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

In Model 2, we found that the effect estimate of average shot scale remained stable, and the covariates had no significant effect on TOM. In addition, model fit as assessed by AIC (Akaike information criterion) deteriorated slightly when including the additional covariates (726.0 vs. 733.3) suggesting that Model 1 produces more precise effect estimates and predictions of TOM.

In a separate analysis, we additionally controlled for self-reported gender and age in both models. The fit of these models was slightly worse and parameter test results were insignificant, which suggests that TOM does not vary across participants with different gender and age. We also tested whether the variables moderated the observed treatment effect of average shot scale, which was not the case.

Discussion

This study examined the effect of average shot scale on TOM responses in film viewers. Confirming our main hypothesis, average shot scale significantly predicted the occurrence of TOM responses in viewers' film story descriptions. A negative association was found between average shot scale and TOM, where decreasing average shot scale (increasing proximity) was associated with increasing TOM responding. The results also suggest that film duration influenced TOM, with longer duration leading to more TOM responses.

The findings of this study provide support for the previous assertion that formal features affect the emotional and cognitive processing of narratives, and it specifically

confirms that not only presentational, but formal features of spatial distance, that is, changes in average shot scale, can impact TOM in viewers. This is line with the findings of previous researchers, who suggested that image size, viewing distance, and shot scale play a role in viewers' involvement with characters. The results were consistent with the argument that spatial proximity increases the motivational relevance of an image. Following the LC4MP model, we assumed that increased motivational relevance increases the allocation of additional resources that may raise the probability that the characters' mental states become part of the viewer's mental representation. The association between the proportion of TOM responses exhibited by viewers and spatial distance suggests that spatial proximity can influence the extent to which characters' mental states are stored in the mental representation of the narrative. To put it differently, higher spatial proximity brings the domain of characters' consciousness into the foreground of the narrative experience (Bruner, 1986).

In addition, a somewhat unexpected finding was that TOM responding is independent of arousal and identification in this model. As can be seen in Model 2, neither of the included covariates of the film experience affected TOM responding. TOM is considered to be an important mechanism in perspective taking that is a subdimension of identification (see introductory section). Thus it could have been assumed that identification would positively influence TOM count, even though it was only a control variable. The lack of connection between variables may be explained by the different type of measurements used for assessing identification and TOM. Identification was determined by self-report measure, whereas TOM was calculated by a performance measure. Another possible explanation is that perspective taking as measured by identification and TOM as measured by mental state references in film descriptions tackle different constructs.

Arousal was also found to have no significant effect on TOM responding in this model. It indicates that the emotional intensity of the experience does not influence viewers' awareness of mental states in fictional characters. As described in the introductory section, arousal and TOM can be considered as ingredients in affective and cognitive empathy, respectively. The lack of effect of arousal on TOM responding is in line with theory and prior research claiming that affective empathy and cognitive empathy are independent processes (Lieberman, 2007; Raz & Hendler, 2014; Wallentin, Simonsen, & Nielsen, 2013).

The results also suggest that TOM response is not only a trait-like capacity, but rather also dependent on the visualization strategy of a narrative. In other words, directors can regulate the extent to which audiences ascribe mental states to characters regardless of viewers' general tendency to read minds. This is in line with previous findings (Bálint, Nagy, & Csabai, 2014; Meins et al., 2014). The significance of this notion comes from the fact that observed out-group members compared with in-group members trigger fewer and less complex TOM responses from an observer (Adams et al., 2010; Cortes, 2005). This is a crucial element in dehumanization processes and leads to deficits in cross-cultural communication. It is possible that visualization techniques could circumvent this bias.

The observation that statistical average shot scale impacts audiences' perception of characters may extend to a broad range of media effects that involve character engagement (e.g., narrative persuasion, enjoyment). Previous research has already demonstrated the importance of shot scale in charity donation (Cao, 2013) and in evaluations of politicians' credibility (Mutz, 2007). We believe that entertainment–education, a campaign strategy often exploiting narrative effects carried by narrative engagement (Moyer-Gusé, 2008), can benefit from the implications of this study as well. Another important implication of this line of research could be the application of the results into visual tools for the improvement of TOM ability. However, in order to do that, we first need a comprehensive model of visual determinants of TOM. Determining the visual procedures of mediated messages in delivering TOM relevant cues can help to strengthen the TOM-enhancing potential of mediated messages. This study represents a first step in this direction.

Limitations

The limitations associated with the methodology of this study have to be acknowledged. Investigating the impact of formal features on psychological responses in viewers of film narratives is one of the most challenging tasks in media psychology. One of the reasons why this task is often left untouched is the technical and methodological difficulty in creating true experiments that explore these connections. Formal features are often difficult to manipulate in an already existing visual narrative, as the quality of the message is easily compromised. On the other hand, creating original visual narratives in different versions is a very costly procedure or is not possible at all, and sufficient quality is not guaranteed. Owing to these considerations, in the first step we decided to sample different messages to represent unique treatment levels in average shot scale. However, this decision raises some methodological concerns. The main limitation of such a quasi-experimental design is that it is almost impossible to control for all the potential confounding variables and to disentangle their potential influences in the statistical model. A systematic selection of the message as in the current study can minimize the specific confounds; however, it is inevitable that there are

still some unobserved variables that have to be controlled. For the present study those variables that covary with average shot scale and also have an influence on TOM are of special importance (Canini, Benini, & Leonardi, 2013). Relatedly, animation films compared with live action films present story events at a higher level of abstraction. The visual presentation of the fictional world is less detailed and many of the form attributes (e.g., the color of the background) are kept constant in animations, which may reduce the number of potential confounding variables. In addition, to control the effect of unobserved formal features, we assessed many film-related responses (e.g., skin conductance, enjoyment, narrative understanding) and introduced them as control variables to the regression model. We also included the duration of movies as control variables. Future research should attempt to measure further movie characteristics in a larger sample of movies to strengthen the findings of the present study.

Conclusion

The visualization of fictional characters plays a role in audiences' narrative processing. The data demonstrate that average shot scale of character visualization influences viewers' awareness of mental states in fictional characters. Perceived spatial distance to a character in a movie brings characters' mental states into the foreground resulting in more mentalized mental models of the narrative. This study is a step toward to the understanding of the relationship between formal features and higher complexity film responses. Knowing how to elicit TOM through message design holds great promise for gaining valuable insight into the mechanism of narrative effects mediated by character engagement.

References

Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., … Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience, 22*, 97–108. doi: 10.1162/jocn.2009.21187

Agresti, A. (2003). *Categorical data analysis*. Hoboken, NJ: Wiley.

Balazs, B., & Carter, E. (2013). *Bela Balazs: Early film theory: Visible man and the spirit of film*. New York, NY: Berghahn Books.

Bálint, K., Nagy, T., & Csabai, M. (2014). The effect of patient-centeredness and gender of professional role models on trainees' mentalization responses. Implications for film-aided education. *Patient Education and Counseling, 97*, 52–58. doi: 10.1016/j.pec.2014.06.005

Baranowski, A. M., & Hecht, H. (2014). The big picture: Effects of surround on immersion and size perception. *Perception, 43*, 1061–1070. doi: 10.1068/p7663

Barnes, J. L., Lombardo, M. V., Wheelwright, S., & Baron-Cohen, S. (2009). Moral dilemmas film task: A study of spontaneous narratives by individuals with autism spectrum conditions. *Autism Research, 2*(3), 148–156.

Baron-Cohen, S. (2001). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The reading the mind in the eyes test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 241–251. doi: 10.1111/1469-7610.00715

Bellman, S., Schweda, A., & Varan, D. (2009). Viewing angle matters – screen type does not. *Journal of Communication, 59*, 609–634. doi: 10.1111/j.1460-2466.2009.01441.x

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91.

Black, J., & Barnes, J. L. (2015). Fiction and social cognition: The effect of viewing award-winning television dramas on theory of mind. *Psychology of Aesthetics, Creativity, and the Arts, 9*, 423–429. doi: 10.1037/aca0000031

Boyarskiy, A. (Producer), Selyanov, S. (Producer), & Bronzit, K. (Director). (2007). *Ubornaya istoriya* (Lavatory Lovestory) [Motion picture]. Russia: Melnitsa Animation Studio.

Bowen, C. J., & Thompson, R. (2013). *Grammar of the shot*. Boston, MA: Focal Press.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.

Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology, 12*, 321–347. doi: 10.1080/15213260903287259

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia, 40*(8), 1129–1138.

Canini, L., Benini, S., & Leonardi, R. (2011). Affective analysis on patterns of shot types in movies. *Image and Signal Processing and Analysis (ISPA), Conference Publications*, 253–258.

Canini, L., Benini, S., & Leonardi, R. (2013). Classifying cinematographic shot types. *Multimedia Tools and Applications, 62*, 51–73. doi: 10.1007/s11042-011-0916-9

Cao, X. (2013). The effects of facial close-ups and viewers' sex on empathy and intentions to help people in need. *Mass Communication and Society, 16*(2), 161–178.

Carroll, N. (1993). Toward a theory of point-of-view editing: Communication, emotion, and the movies. *Poetics Today, 14*(1), 123–141.

Carroll, N., & Seeley, W. P. (2013). Cognitivism, psychology, and neuroscience: Movies as attentional engines. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies* (pp. 53–75). New York, NY: Oxford University Press.

Chung, A. H., & Slater, M. D. (2013). Reducing stigma and outgroup distinctions through perspective-taking in narratives. *Journal of Communication, 63*(5), 894–911.

Codispoti, M., & De Cesarei, A. (2007). Arousal and attention: Picture size and emotional reactions. *Psychophysiology, 44*, 680–686. doi: 10.1111/j.1469-8986.2007.00545.x

Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society, 4*(3), 245–264.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of

mind" in people with schizophrenia. *Schizophrenia Research, 17*(1), 5–13.

Cortes, B. P. (2005). Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality and Social Psychology Bulletin, 31*, 243–253. doi: 10.1177/0146167204271421

De Cesarei, A., & Codispoti, M. (2006). When does size not matter? Effects of stimulus size on affective modulation. *Psychophysiology, 43*(2), 207–215.

De Cesarei, A., & Codispoti, M. (2008). Fuzzy picture processing: Effects of size reduction and blurring on emotional processing. *Emotion, 8*, 352–363. doi: 10.1037/1528-3542.8.3.352

De Cesarei, A., & Codispoti, M. (2010). Effects of picture size reduction and blurring on emotional engagement. *PloS One, 5*(10), e13399.

Decety, J., & Ickes, W. J. (2009). *The social neuroscience of empathy*. Cambridge, MA: MIT Press.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71–100.

Detenber, B. H., & Lang, A. (2010). The influence of form and presentation attributes of media on emotion. In K. Doveling, C. von Scheve, & E. A. Konijn (Eds.), *The Routledge handbook of emotions and mass media* (pp. 275–293). New York, NY: Routledge.

Detenber, B. H., & Reeves, B. (1996). A bio-informational theory of emotion: Motion and image size effects on viewers. *Journal of Communication, 46*(3), 66–84.

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PLoS ONE, 8*, e81279. doi: 10.1371/journal.pone.0081279

de Wit, M. D. (Director). (2001). *Father and Daughter [Motion picture]*. The Netherlands: CinéTé Filmproductie BV.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., … Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36*(5), 623–636.

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the multifaceted empathy test (MET). *Journal of Autism and Developmental Disorders, 38*, 464–473. doi: 10.1007/s10803-007-0486-x

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 498–541). Hoboken, NJ: Wiley.

Fonagy, P., & Bateman, A. (2007). Mentalizing and borderline personality disorder. *Journal of Mental Health, 16*(1), 83–101.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999–1010.

Freelon, G. D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science, 5*(1), 20–33.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*, 694–724. doi: 10.1037/0033-2909.133.4.694

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind". *Trends in Cognitive Sciences, 7*, 77–83. doi: 10.1016/S1364-6613(02)00025-6

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The "Reading the Mind in Films" Task: Complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience, 1*, 111–123. doi: 10.1080/17470910600980986

Hall, E. T., & Hall, E. T. (1969). *The hidden dimension*. New York, NY: Anchor Books.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252–264.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders, 30*(3), 225–236.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091.

Hood, B.M., Macrae, C.N., Cole-Davies, V., & Dias, M. (2003). Eye remember you: The effects of gaze direction on face recognition in children and adults. *Developmental Science, 6*, 67–71. doi: 10.1111/1467-7687.00256

Hou, J., Nam, Y., Peng, W., & Lee, K. M. (2012). Effects of screen size, viewing angle, and players' immersion tendencies on game experience. *Computers in Human Behavior, 28*, 617–623. doi: 10.1016/j.chb.2011.11.007

Howard, I. P. (2012). *Perceiving in depth*. New York, NY: Oxford University Press.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews, 33*, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia, 45*(5), 1019–1028.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*(6156), 377–380.

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology, 90*, 543–555. doi: 10.1037/0022-3514.90.4.543

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication, 50*, 46–70. doi: 10.1111/j.1460-2466.2000.tb02833.x

Lang, A., Bradley, S. D., Park, B., Shin, M., & Chung, Y. (2006). Parsing the resource pie: Using STRTs to measure attention to mediated messages. *Media Psychology, 8*(4), 369–394.

Levin, D. T., Hymel, A. M., & Baker, L. (2013). Belief, desire, action, and other stuff: theory of mind in movies. In A. P. Shimamura (Ed.), *Psychocinematics* (pp. 244–266). New York, NY: Oxford University Press.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58*, 259–289.

Lombard, M. (1995). Direct responses to people on the screen: Television and personal space. *Communication Research, 22*, 288–324. doi: 10.1177/009365095022003002

Lombard, M., Ditton, T. B., Grabe, M. E., & Reich, R. D. (1997). The role of screen size in viewer responses to television fare. *Communication Reports, 10*(1), 95–106.

Lombard, M., Reich, R. D., Grabe, M. E., Bracken, C. C., & Ditton, T. B. (2000). Presence and television. *Human Communication Research, 26*(1), 75–98.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103–134. doi: 10.1146/annurev-psych-120709-145406

Meins, E., & Fernyhough, C. (2010). *Mind-mindedness coding manual, Version 2.0 (Unpublished manuscript)*. Durham, UK: Durham University.

Meins, E., Fernyhough, C., & Harris-Waller, J. (2014). Is mind-mindedness trait-like or a quality of close relationships? Evidence from descriptions of significant others, famous people, and works of art. *Cognition, 130*(3), 417–427.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an "intentional stance" on eye-gaze shifts: A functional neuroimaging study of social perception in children. *NeuroImage, 27*(1), 247–252.

Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: explaining the persuasive effects of entertainment-education messages. *Communication Theory, 18*, 407–425. doi: 10.1111/j.1468-2885.2008.00328.x

Mühlberger, A., Neumann, R., Wieser, M. J., & Pauli, P. (2008). The impact of changes in spatial distance on emotional responses. *Emotion, 8*(2), 192.

Mutz, D. C. (2007). Effects of "in-your-face" television discourse on perceptions of a legitimate opposition. *American Political Science Review, 101*(04), 621–635.

Persson, P. (2003). *Understanding cinema: A psychological theory of moving imagery*. Cambridge and New York, MA: Cambridge University Press.

Plantinga, C. R. (1999). The scene of empathy and the human face on film. In C. R. Plantinga & G. M. Smith (Eds.), *Passionate views: Film, cognition, and emotion* (pp. 239–255). Baltimore, MD: Johns Hopkins University Press.

Raz, G., & Hendler, T. (2014). Forking cinematic paths to the self: Neurocinematically informed model of empathy in motion pictures. *Projections, 8*(2), 89–114.

Reeves, B., & Geiger, S. (1994). Designing experiments that assess psychological responses to media messages. In A. Lang (Ed.), *Measuring psychological responses to media* (pp. 165–180). Hillsdale, NJ: Lawrence Erlbaum.

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology, 1*(1), 49–67.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge and New York, MA: Cambridge University Press.

Salt, B. (1992). *Film style and technology: History and analysis*. London, UK: Starword.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*(3), 617–627.

Shuskhov, A. (Director). (2010). *Invention of Love [Motion picture]*. Russia: HHG Film Company.

Smith, M. S. (1995). *Engaging characters: Fiction, emotion, and the cinema*. Oxford, UK: Clarendon Press.

Tal-Or, N., & Cohen, J. (2010). Understanding audience involvement: Conceptualizing and manipulating identification and transportation. *Poetics, 38*, 402–418. doi: 10.1016/j.poetic.2010.05.004

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., … Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage, 29*(1), 90–98.

Wallentin, M., Simonsen, A., & Nielsen, A. H. (2013). Action speaks louder than words: Empathy mainly modulates emotions from theory of mind-laden parts of a story. *Scientific Study of Literature, 3*, 137–153. doi: 10.1075/ssol.3.1.11wal

Whitaker, M. (Director). (2008). *Lettin'go. [Motion picture]*. France: Living Heart.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292–297.

Received August 28, 2015
Revision received December 14, 2015
Accepted February 6, 2016
Published online August 26, 2016

Katalin Bálint
Department of Media, Knowledge and Communication
University of Augsburg
Germany
katalin.balint@phil.uni-augsburg.de

Katalin Bálint is a postdoc researcher at the Institute of Media, Knowledge, and Communication, University of Augsburg, Germany. She has a background in psychology and film and communication studies. Her research interest is centered on the psychological processes underlying audience members' responses to media and the role of audiovisual formal features in information processing.

Thomas Klausch is a statistician at the Department of Epidemiology and Biostatistics at VU University Medical Center Amsterdam, The Netherlands. Before joining VU he was a postdoc and PhD student at the Department of Methodology and Statistics at Utrecht University and the Dutch National Office of Statistics, Statistics Netherlands.

Tibor Pólya (PhD, 2003) is a research fellow affiliated at the Institute of Cognitive Neuroscience and Psychology and an associate professor at the Károli University of the Reformed Church, Hungary. His interests are the narrative organization of emotional experiences and the role of group history narratives in defining social identities.

